

WHAT IS CLAIMED IS:

1. A videophone device for transmitting/receiving an image and voice to/from another device through a network, comprising:

5 a voice input unit configured to input voice data;
 an image input unit configured to input image data;

 a text data generating unit configured to generate text data while at least one of the image data and the voice data is being input;

10 a synthesizing unit configured to synthesize the voice data, the image data and the text data to obtain data; and

 a communication unit configured to transmit the data obtained by the synthesizing unit.

15 2. The videophone device according to claim 1, wherein the synthesizing unit generates relevant information indicating a relationship of the text data with the image data and the voice data with respect to time.

20 3. The videophone device according to claim 1, wherein the text data generating unit includes a voice recognizing unit configured to execute voice recognition on the voice data input by the voice input unit, to thereby generate text data.

25 4. The videophone device according to claim 1, wherein the text data generating unit includes a text

data input unit configured to generate text data based on the data input from an input device.

5 5. The videophone device according to claim 1,
wherein the synthesizing unit includes an adjusting
unit configured to adjust synthesizing of the text data
generated by the text data generating unit with the
image data and the voice data, such that reproduction
of the image and the voice by the other device is
synchronized with displaying of the text by the other
10 device.

 6. The videophone device according to claim 5,
wherein the adjusting unit is configured to adjust
a displaying time period such that a text based on the
text data is displayed for a longer time period than
15 that for which voice is input by the voice input unit.

 7. A videophone device for transmitting/receiving
an image and voice to/from another device through
a network, comprising:

 a communication unit configured to receive,
20 through a network, data in which image data and text
data are synthesized;

 a dividing unit configured to divide the data
received by the communication unit into the image data
and the text data;

25 an image processing unit configured to synthesize
a text based on the text data obtained by dividing of
the dividing unit with the image data obtained by

dividing of the dividing unit; and

an image output unit configured to output an image based on the image data with which the text is synthesized by the image processing unit.

5 8. The videophone device according to claim 7, further comprising:

a storage unit configured to store the data received by the communication unit; and

10 a recording/reproducing unit configured to cause the data stored in the storage unit to be divided by the dividing unit.

15 9. The videophone device according to claim 7, further comprising an adjusting unit for adjusting a timing at which the text is synthesized with the image data by the image processing unit.

10. A videophone device which is to be connected to another device through a network, comprising:

an image input unit configured to input image data;

20 a text data input unit configured to input text data while the image data is being input by the image input unit;

25 a synthesizing unit configured to synthesize the image data and the text data to obtain synthetic data; and

a communication unit configured to transmit the synthetic data obtained by the synthesizing unit,

through the network.

11. A videophone device which is to be connected to another device, through a network, comprising:

5 a communication unit configured to receive data, in which image data and text data are synthesized, through the network;

a dividing unit configured to divide the data received by the communication unit into the image data and the text data;

10 a voice synthesizing unit configured to perform voice synthesis based on the text data obtained by dividing of the dividing unit;

15 a voice output unit configured to output synthetic voice obtained by the voice synthesis performed by the voice synthesizing unit; and

an image output unit configured to output an image based on the image data obtained by the dividing of the dividing unit.

20 12. A videophone device configured to transmit/receive an image and voice to/from another device through a network, comprising:

a information receiving unit configured to receive information indicating a unit provided in the other device, from the other device, through the network;

25 a voice input unit configured to input voice data; an image input unit configured to input image data;

a text data generating unit configured to generate text data while the image data and the voice data are being input by the image input unit and the voice input unit, respectively;

5 a synthesizing unit configured to selectively synthesize the voice data, the image data and the text data in accordance with the information indicating the unit provided in the other device, which is received by the information receiving unit, thereby obtaining
10 synthetic data; and

a transmitting unit configured to transmit the synthetic data obtained by the synthesizing unit, through the network.

13. The videophone device according to claim 12,
15 further comprising:

an information transmitting unit configured to transmit information indicating the units provided in the videophone device, to the other device, through the network; and

20 a setting unit configured to set the units in accordance with the information transmitted by the information transmitting unit, in such a manner as to allow an optional one or ones of the units to be used.

14. A data transmitting/receiving method of a
25 videophone device for transmitting/receiving an image and voice to/from another device through a network, comprising:

generating first voice data and first image data,
and generating first text data while inputting the
first image data and the first voice data;

5 synthesizing the first voice data, the first image
data and the first text data to obtain synthetic data,
and transmitting the synthetic data;

receiving data transmitted from the other device
through the network;

10 dividing the received data into second image data
and second text data; and

adding the second text data to the second image
data to obtain synthetic data.

15 15. The method according to claim 14, further
comprising executing voice recognition on the first
voice data, to thereby obtain the first text data.

20 16. The method according to claim 14, further
comprising adjusting synthesizing of the first text
data with the first image data and the first voice data
such that reproduction of an image and voice by the
other device is synchronized with displaying of a text
by the other device.

25 17. A data transmitting/receiving method of
a video phone system for transmitting/receiving
an image and voice to/from a videophone device through
a network, comprising:

in a first videophone device, (i) inputting voice
data and image data, and generating text data while

inputting the voice data and the image data; and (ii) synthesizing the voice data, the image data and the text data to obtain synthetic data, and transmitting the synthetic data through the network, and

5 in a second vide phone device, (i) receiving the data transmitted from the first videophone device through the network, (ii) dividing the image data and the text data of the transmitted data , and (iii) synthesizing a text based on the text data with the
10 image data to obtain synthetic data, and outputting the synthetic data.

18. The method according to claim 17, further comprising adjusting synthesizing of the text data generated by the text data generating unit with the
15 image data and the voice data such that reproduction of an image and voice by the other device is synchronized with displaying of a text by the other device.

19. A data transmitting/receiving method of a videophone device for transmitting/receiving an image
20 and voice to/from another videophone device through a network, comprising:

 in a first videophone device, (i) inputting image data, and inputting text data while inputting the image data; and (ii) synthesizing the image data and the text
25 data to obtain synthetic data, and outputting the synthetic data, and

 in a second videophone device, (i) receiving the

synthetic data transmitted from the first videophone device, through the network, (ii) dividing the transmitted data into the image data and the text data, and (iii) performing voice synthesis based on the text data to output voice, and output an image based on the image data.